Draft version May 13, 2024
Typeset using LaTeX twocolumn style in AASTeX631

A Golden Sample with precise Atmospheric Parameter labels of M-type Stars

Bing Du [1], A-Li Luo [1,2], Song Wang [1], Yinbi Li [1], Cai-Xia, Qu [1,2] Xiao Kong [1], Yan-xin Guo [1], Yi-han Song [1], and Fang Zuo [1]

[1] *CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, China*
[2] *School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*

ABSTRACT

The discrepancies between the theoretical and observed spectra, and the systematic differences between the spectroscopic measurements make the measurement of atmospheric parameters of M-type stars complicated. In this work, we present a golden sample with precise atmospheric parameter labels of M-type stars through stellar label transfer and sample cleaning. We addressed systematic discrepancies in spectroscopic measurements employing StarHorse as the reference standard. We used density-dased spatial clustering of applications with noise (DBSCAN) to remove the unreliable samples in each subgrid of parameters. To confirm the reliability of the stellar labels, a 5-layer neural network was utilized, randomly partitioning the samples into training and testing sets. The testing set shows variations of 14 K, 0.06 dex, and 0.05 dex in $T_{\rm eff}$, $\log g$, and [M/H], respectively. In addition, we conducted an internal cross-validation to enhance the validation. We obtained precisions of 11 K, 0.05 dex, and 0.05 dex for $T_{\rm eff}$, $\log g$, and [M/H], respectively, suggesting that the annotations of the parameters in the golden sample are precise. The golden sample was utilized in the LAMOST stellar parameter pipeline for M-Type stars (LASPM), producing an almost seamless Kiel distribution diagram for LAMOST DR11 data. The updated LASPM shows improved precision compared to its predecessor, for S/N higher than 10, with improvements from 118 to 67 K in $T_{\rm eff}$, 0.2 to 0.07 dex in $\log g$, and 0.29 to 0.14 dex in [M/H].

Keywords: techniques, spectroscopic – methods, data analysis–methods: statistical

1. INTRODUCTION

The atmospheric parameters of M-type stars offer valuable information on revealing the formation history of the Galaxy. The M dwarf stars, which dominate the faint magnitudes of the Galaxy (Bochanski et al. 2010), are not only important for determining the initial mass function but also good tracers of the chemical and dynamical history of the Milky Way because of their exceptionally long lifetime. The M Giants with high luminosity are good tracers for revealing the accretion and merger events in the Galaxy by discovering and identifying substructures in the Galactic outer disk and remnants of stellar streams in the halo (Li et al. 2023). However, stellar parameters of M-type stars such as effective temperature (Teff), surface gravity (log g), and metallicity ([M/H]) are model dependent.

Impressive progress has been made in the study of atmospheric models and molecular absorption of late-type stars in the three decades. For example, the PHOENIX BT-Settl model, using revised solar abundances and updated atomic and molecular line opacities (Allard et al. 2012, 2013), can reproduce the observed spectra very well for M dwarfs (Rajpurohit et al. 2013). Consequently, the BT-Settl model grids were utilized by Du et al. (2021) to establish the LAMOST stellar parameter pipeline for M-type stars (LASPM). However, the BT-Settl model has a poor fit to the observed spectra for M giants, making the LASPM parameters of M giants unreliable (Du et al. 2021; Qiu et al. 2023). Compared with the BT-Settl model, the MARCS model has a good fit with both observations and synthetic spectral energy distributions (SEDs) for M giants from M1 to M6III, but it has an excess blue flux for M dwarfs (Gustafsson

Corresponding author: A-Li Luo, Song Wang
lal@nao.cas.cn, songw@bao.ac.cn

et al. 2008; Plez 2008). Moreover, it is still difficult to obtain a consistent metallicity of M stars in the optical region from both models (Passegger et al. 2016; Du et al. 2021).

In addition to the discrepancies between the theoretical and observed spectra, there are systematic errors between the spectroscopic measurements. The LASPM overestimates $\log g$ by 0.63 dex and underestimates [M/H] by 0.25 dex when compared to the catalog of the Sloan Digital Sky Survey Apache Point Observatory Galactic Evolution Experiment (SDSS/APOGEE) (Du et al. 2021). The revised LASPM, including Data Release 9 (DR9) and subsequent, continues to overestimate $\log g$ by 0.27 dex relative to APOGEE. Ding et al. (2022) determined the parameters of M-type stars of LAMOST DR8 by applying the MILES interpolator to the ULySS package. Systematic errors still exist between the results of Ding et al. (2022) and those of the APOGEE Stellar Parameter and Chemical Abundances Pipeline (ASPCAP). The causes of systematic errors between spectroscopic measurements are unclear.

Thanks to efforts to decode the stellar parameters from high-resolution spectra (Rajpurohit et al. 2014; Passegger et al. 2016; Veyette et al. 2017; Rajpurohit et al. 2018), the ASPCAP of SDSS Data Release 16 (DR16) determined effective temperatures down to 3000 K by using new atmospheric grids (Jönsson et al. 2020). The parameters of late-type stars from high-resolution spectra enables the data-driven methods to measure the stellar parameters of M-type stars from either low-resolution spectra or multi-band photometric data by transferring stellar labels. Li et al. (2021) determined $\sim$ 300,000 spectroscopic stellar parameters ($T_{\rm eff}$ and [M/H]) of M dwarfs by training a Stellar LAbel Machine (SLAM) model using the LAMOST spectra with APOGEE DR16 stellar labels. Employing the SLAM model, Qiu et al. (2023) determined the stellar parameters ($T_{\rm eff}$, $\log g$, [M/H], α/M]) for over 43,000 M giants by utilizing the LAMOST spectra and APOGEE DR17 stellar labels. Building on the previous research, Qu et al. (2024) obtained atmospheric parameters for 1,806,921 cool dwarfs using Gaia DR3, employing machine learning algorithms trained on multiband photometry and stellar labels from APOGEE DR16, along with catalogs from Li et al. (2021) and Ding et al. (2022).

In addition to the spectroscopic measurements, the StarHorse code calculated the photo-astrometric stellar parameters from Gaia data combined with the photometric catalogs of Pan-STARRS1, SkyMapper, 2MASS, and AllWISE (Anders et al. 2019, 2022). Thanks to the higher precision of the Gaia EDR3 and the new stellar-density priors of StarHorse, the precision over previous estimates was substantially improved with a typical precision of 3% (15%) in distance and 140 K (180 K) in $T_{\rm eff}$ at magnitude G = 14 (17) (Anders et al. 2022). Although the photo-astrometric parameters are not as accurate as the spectroscopic measurements, the statistical distribution of these parameters can serve as a reference standard for adjusting the systematic discrepancies found in the spectroscopic measurements.

By gathering parameter measurements from LAMOST late-type stars and using the StarHorse catalog as a standard, it is possible to filter out a golden sample with precise atmospheric parameter labels. The golden sample, characterized by a balanced parameter distribution, can serve as an empirical library for the determination of stellar parameters and also provides data annotations for machine learning algorithms driven by data. By employing this golden sample as reference stars, the LASPM can circumvent the issue of mismatches between theoretical and observed spectra.

In this work, we present a golden sample with precise stellar labels of M-type stars through stellar label transfer and sample cleaning. We address systematic discrepancies in spectroscopic measurements employing StarHorse as the reference standard. Given the dense forest of spectral features of M-type stars in the optical band, we select clean samples in each parameter bin rather than co-adding spectra as outlined in Du et al. (2019). We use density-dased spatial clustering of applications with noise (DBSCAN) to remove the unreliable samples in each subgrid of parameters. The golden sample is employed in LASPM for the determination of stellar parameters of M-type stars, providing data products for the upcoming LAMOST data release (DR11 and subsequent). This paper is organized as follows. In Section 2, we detail sample selection and cleaning. In Section 3, we demonstrate the verification of the stellar labels for our samples and the results of this golden sample, including the stellar parameter coverage and the spectra. In Section 4, we introduce the application of the golden sample to LASPM. Finally, the main conclusions are summarized in Section 5.

2. SAMPLE SELECTION AND CLEANING

2.1. *Primary Samples*

The LAMOST low-resolution survey has collected more than 10 million spectra (R $\sim$ 1800, 3800–9000 Å). The spectra were analyzed by the LAMOST 1D pipeline to recognize their spectral classes and determine the radial velocity (RV) for stars (Luo et al. 2015). For LAMOST DR11, the LAMOST 1D pipeline recognized more than 1 million M-type stars with spectral types of M1–M9 and roughly luminosity classes of M giants (gM) and M dwarfs (dM). Since the spectral features of cool stars are dominant in the i-band, we selected the LAMOST M-type samples excluding the stars with a spectral Signal-to-Noise ratio (S/N) less than 10 in the i-band. Unless otherwise indicated, all the S/Ns presented hereafter refer to the S/N_i value.

We selected primary samples based on their locations in the absolute magnitude G versus the BP - RP diagram by cross-matching the LAMOST M-type catalog with Gaia EDR3 (Gaia Collaboration et al. 2021). We excluded unreliable photometry and unresolved binary stars using the same criteria described in Qu et al. (2024). We excluded variable stars by cross-matching our sample with a list of variable stars collected from various surveys, including Kepler (McQuillan et al. 2013, 2014; Kirk et al. 2016; Santos et al. 2019), ZTF (Chen et al. 2020), K2 (Reinhold & Hekker 2020), WISE (Chen et al. 2018), Gaia (Gaia Collaboration et al. 2023), TESS (Howard et al. 2022; Prsa et al. 2022), and LAMOST (Xu et al. 2022). We adopted the photogeometric distances of (Bailer-Jones et al. 2021) as the distances of our sample, which was inferred from EDR3 parallaxes with a zero-point correction based on a three-dimensional Galactic model. The 3D dust map method (Green et al. 2018) was employed to correct the extinction.

2.1.1. *M Dwarfs*

For M dwarfs, considering only those close to us can be observed because of their low luminosities, we selected our sample with distances less than 2.0 kpc. The G magnitude limits for M dwarfs are set between 7-12 according to the contour shape in the color–magnitude diagram presented in Figure 1. To derive suitable dwarf samples, we calculated the average values and the 1σ uncertainties of G magnitudes using a color step of 0.5 mag. Subsequently, we performed a linear fit to the average values of G magnitudes, indicated by the black dashed line in Figure 1. We shifted the black dashed line up and down by 1σ uncertainty (the average value of 1σ uncertainties), which is 0.8 mag for dwarfs. We obtained the dwarf sample based on the red dashed lines in the color–magnitude diagram shown in Figure 1.

2.1.2. *M Giants*

To maintain the accuracy in the distance measurements for M giants, we excluded objects with distances greater than 5.0 kpc. We set color limits for M giants as BP-RP between 1.25-4.0 according to the contour shape in the color–magnitude diagram presented in Figure 2. For giants, we also calculated the average values and the 1σ uncertainties of G magnitudes with a color step

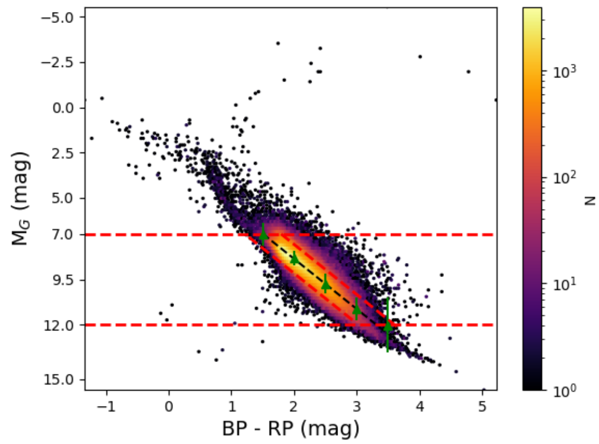

Figure 1. Color-magnitude diagram of the primary sample of M dwarfs. The error-bars represent the average values and the 1σ uncertainties of G magnitudes with a color step of 0.5 mag, and the black dashed line is the first-order polynomial fit to the average values. The oblique red dashed lines are created by shifting the black line up and down by 1σ uncertainty. The objects lie within the red dashed box are selected.

of 0.5 mag. Then we performed a third-order polynomial fit to the the average values (the black dashed curve in Figure 2). We shifted the black dashed curve up and down by 1σ uncertainty (0.8 mag). We selected giant objects that lie within the red dashed enclosed region in the color-magnitude diagram presented in Figure 2.

2.2. *Stellar Parameter Labels*

We labeled our samples with stellar parameters collected from a list of catalogs, including SDSS DR17 ASPCAP(Jönsson et al. 2020), M giant parameters of Qiu et al. (2023) (hereafter referred to as Qiu23), M dwarf parameters of Ding et al. (2022) (hereafter referred to as Ding22), and M dwarf parameters of LASPM. The systematic discrepancies among these catalogs were adjusted by aligning them with the StarHorse catalog (Anders et al. 2022).

2.2.1. *M Dwarfs*

For M dwarfs, we adopted stellar parameters from ASPCAP DR17, Ding22, and LASPM. Figure 3 shows the differences in their respective comparisons with the StarHorse catalog. We found that there were varying offsets among the three catalogs compared to StarHorse. The ASPCAP overestimated $T_{\rm eff}$ by approximately 87 K, while there was no significant systematic difference for $\log g$ and [M/H]. There was no significant shift between the $T_{\rm eff}$ of LASPM and that of StarHorse, but large offsets for $\log g$ and [M/H]. We shifted the stellar

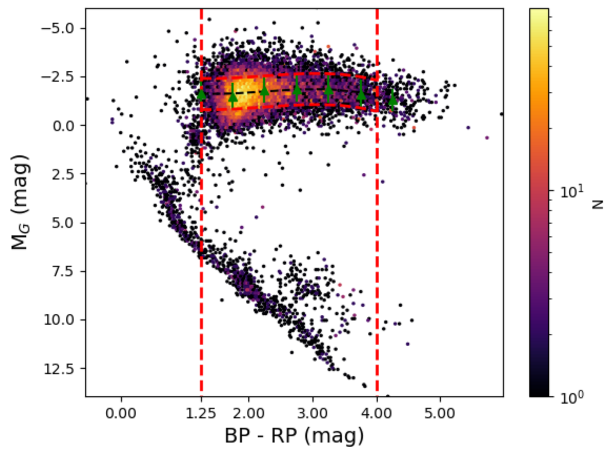

Figure 2. Color-magnitude diagram of the primary sample of M giants. The error-bars represent the average values and the 1σ uncertainties of G magnitudes with a color step of 0.5 mag, and the black dashed curve is the third-order polynomial fit to the average values. The red dashed curves are created by shifting the black curve up and down by 1σ uncertainty. The objects lie within the red dashed enclosed region are selected.

parameters of the three catalogs using their respective systematic offsets relative to StarHorse. We selected dwarf samples with parameter differences of $\Delta T_{\rm eff} < 100$ K, $\Delta \log g < 0.2$ dex, and Δ[M/H] < 0.2 dex given the standard deviation values of the differences presented in Figure 3. In the process of parameter transfer, the ASPCAP catalog was prioritized above Ding22, and Ding22 was prioritized above LASPM. For example, when a star was parameterized by ASPCAP, Ding22, and LASPM, the parameters of ASPCAP were adopted, when parameterized by both Ding22 and LASPM, the parameters of Ding22 were adopted.

2.2.2. *M Giants*

For M giants, we adopted stellar parameters from ASPCAP DR17 and Qiu23. Figure 4 presents the differences between these catalogs and the StarHorse catalog. Since the Qiu23 parameters were transferred from those of ASPCAP, both catalogs have almost the same offsets relative to StarHorse. We still shifted the parameters of both catalogs, using their respective systematic offsets. We selected giant samples with parameter differences of $\Delta T_{\rm eff} < 70$ K, $\Delta \log g < 0.2$ dex, and Δ[M/H] < 0.2 dex given the standard deviation values in Figure 4. In cases where parameter transfer was necessary, the ASPCAP catalog was prioritized over Qiu23. This implies that when a star was cataloged by both ASPCAP and Qiu23, the ASPCAP parameters were adopted.

2.2.3. *Supplement*

Furthermore, to complement the existing samples, we incorporated 160 stars characterized by a LAMOST spectral S/N exceeding 50, along with accurate parameter determinations from ASPCAP ($\Delta T_{\rm eff}< 10$ K, $\Delta \log g< 0.05$ dex, and Δ[M/H] < 0.02 dex). Figure 5 shows their locations on the Gaia color-magnitude diagram, which is color-coded by the log g values. We noted that the giant and dwarf loci correspond to their $\log g$ values and are obvious in the color-magnitude diagram. Due to the high quality of their observed spectra and the precise parameter determinations, the 160 stars were incorporated into our samples. The parameters of the 160 stars were adjusted by using identical corrections from ASPCAP to StarHorse for both dwarf and giant stars.

Currently, we have gathered a sample comprising 45,422 stars, with stellar parameters distributed as follows: 4,191 from ASPCAP, 2,625 from Qiu23, 24,397 from Ding22, and 14,209 from LASPM.

2.3. *Spectral Pre-processing*

2.3.1. *Back to Rest Frames*

For M-type stars, the RV of the LAMOST 1D pipeline was used to shift all of the spectra into their rest frames. The precision quoted for the RV of the LAMOST 1D pipeline was approximately 5.0 km s^{-1}(Luo et al. 2015; Du et al. 2021). The LAMOST low-resolution spectra were sampled in constant-velocity pixels, with a pixel scale of 69 km s^{-1} (Du et al. 2019). The precision of the RV of the LAMOST 1D pipeline (~ 5.0 km s^{-1}) was less than 10% of the pixel scale, which means that its RV calculations are accurate at subpixel values. Therefore, using the RV of the LAMOST 1D pipeline, we completed an accurate shift to the rest frames.

2.3.2. *Flux Calibration and Dereddening*

Before clustering for spectral cleaning, we should correct the dereddening errors associated with the LAMOST flux calibration. As described in Du et al. (2019), the LAMOST flux calibration introduces uncertainties to the SEDs of the calibrated spectra. Following the work done in Du et al. (2019), we recalibrated each spectrum of M-type stars by comparing the observed SED with the synthetic SED, using a second-order multiplicative polynomial to minimize recalibration impacts on the crowded bands of cool stars. For dwarfs, we adopted the BT-Settl model as the reference SED given the blue excess of the MARCS model, while for giants, we opted for the MARCS model over the BT-Settl model considering that the latter has a poor fit with the observed spectra for M giants.

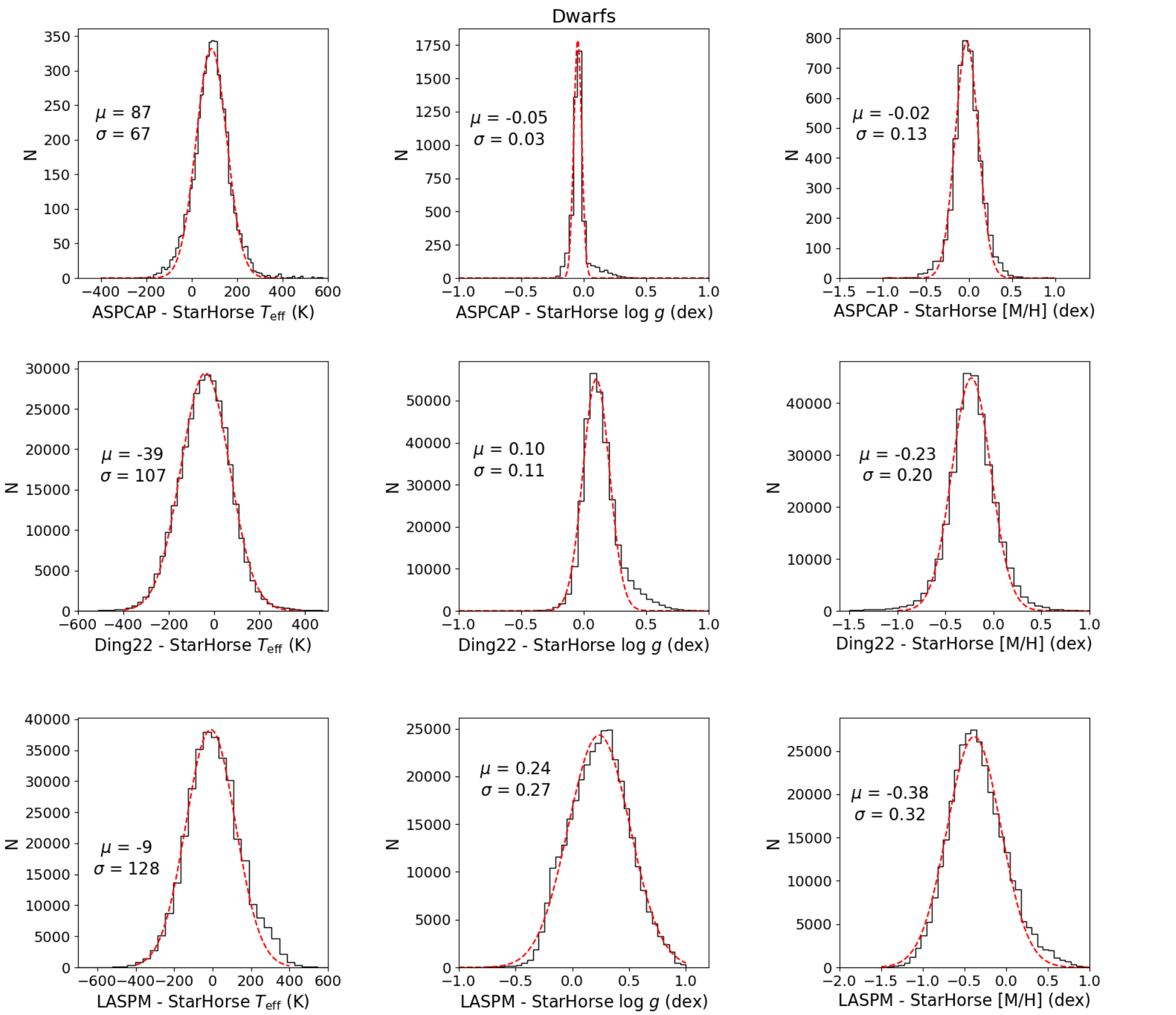

Figure 3. Histograms of differences between ASPCAP and StarHorse (top), Ding22 and StarHorse (middle), and LASPM and StarHorse (bottom). The red dashed curves are Gaussian fits to the distributions, and the mean and dispersion of the Gaussian fits to the mean and standard deviation values of the differences.

2.4. *Sample Cleaning Based on DBSCAN*

After the previous spectral pre-processing, we did spectra clustering in bins of $T_{\rm eff}$, $\log g$, and [M/H] to clean samples. We separated the spectra into the following parameter bins: $T_{\rm eff}$ in steps of 50 K, $\log g$ in steps of 0.1 dex, and [M/H] in steps of 0.1 dex. For groups with more than 150 spectra, only the first 150 spectra with the largest S/Ns were selected for clustering. All of the spectra were resampled to a set of fixed wavelengths to align their wavelengths.

We applied DBSCAN to the spectral clustering in each parameter bin. DBSCAN is robust to noise and can handle clusters of different densities and is also capable of identifying outliers. This makes it a good spectral cleaner for our samples. DBSCAN requires two parameters: eps and minPts, eps determines the radius around each point within which to search for neighboring points, and minPts is the minimum number of points required to form a cluster. Figure 6 displays examples of spectra clustering for eps values of 1.0 and 2.0. We set eps=1.0 for spectral cleaning according to the calculated distances between the optical band spectra with

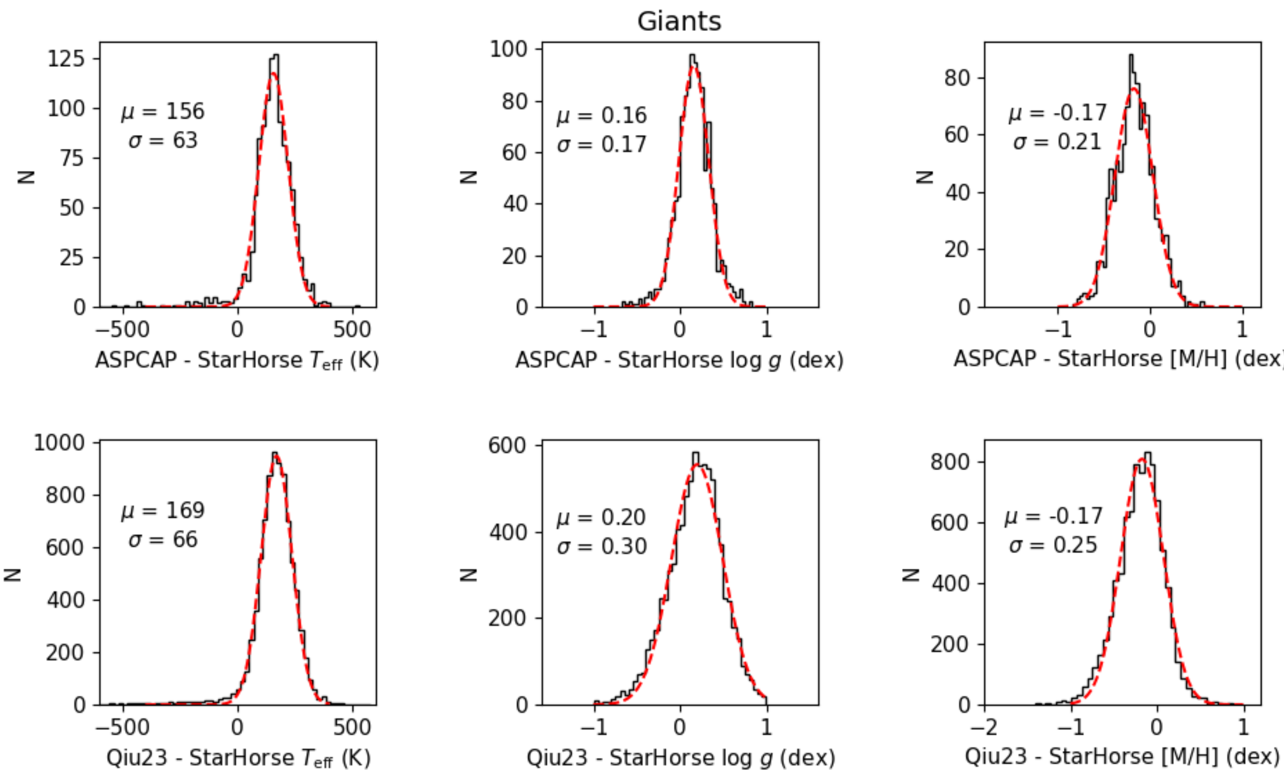

Figure 4. Histograms of differences between ASPCAP and StarHorse (upper), and Qiu23 and StarHorse (lower). The red dashed curves are Gaussian fits to the distributions, and the mean and dispersion of the Gaussian fits to the mean and standard deviation values of the differences.

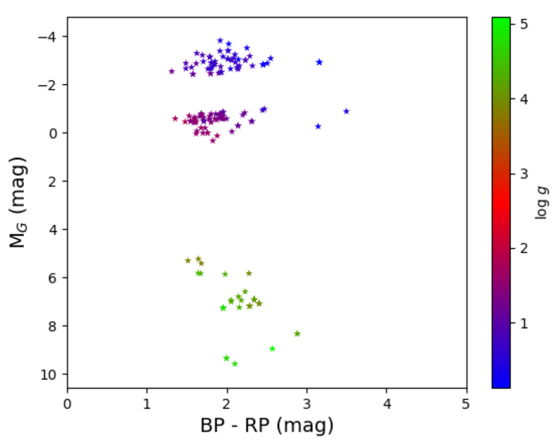

Figure 5. Color-magnitude diagram of the 160 added sample, color-coded by the log g values.

a wavelength coverage of 3900-8800 Å. We set different minPts values from 2 to 5 based on the number of samples in each parameter bin. The sample count in each parameter bin was capped at seven to ensure a balanced dataset suitable for machine learning. For grids that contained only one sample or experienced clustering failure, we manually inspected their spectra for quality control. We eventually obtained 5132 M-type stars after clustering cleaning and manual inspection.

2.5. *Further Cleaning Based on Neural Network*

To further cleanse and validate the self-consistency of stellar labels for the 5132 stars, we trained a 5-layer neural network by randomly separating the 5132 stars into a training set and a testing set. The structure of neural networks consists of a fully connected network with an input layer, three hidden layers, and an output layer. Considering that LASPM only used the red part spectra, we also used the red part spectra (6000-8800 Å) in neural network training. The spectral features were reduced from 1657 to 512, 512 to 64, then the 64 features were mapped to the output layer of 3 stellar labels. We iterated training 5 times and eliminated outliers until the testing set had all points with parameter errors of $\Delta T_{\rm eff} < 100$ K, $\Delta \log g < 0.2$ dex, and Δ[M/H] < 0.2 dex. We removed a total of 27 stars during this clean-

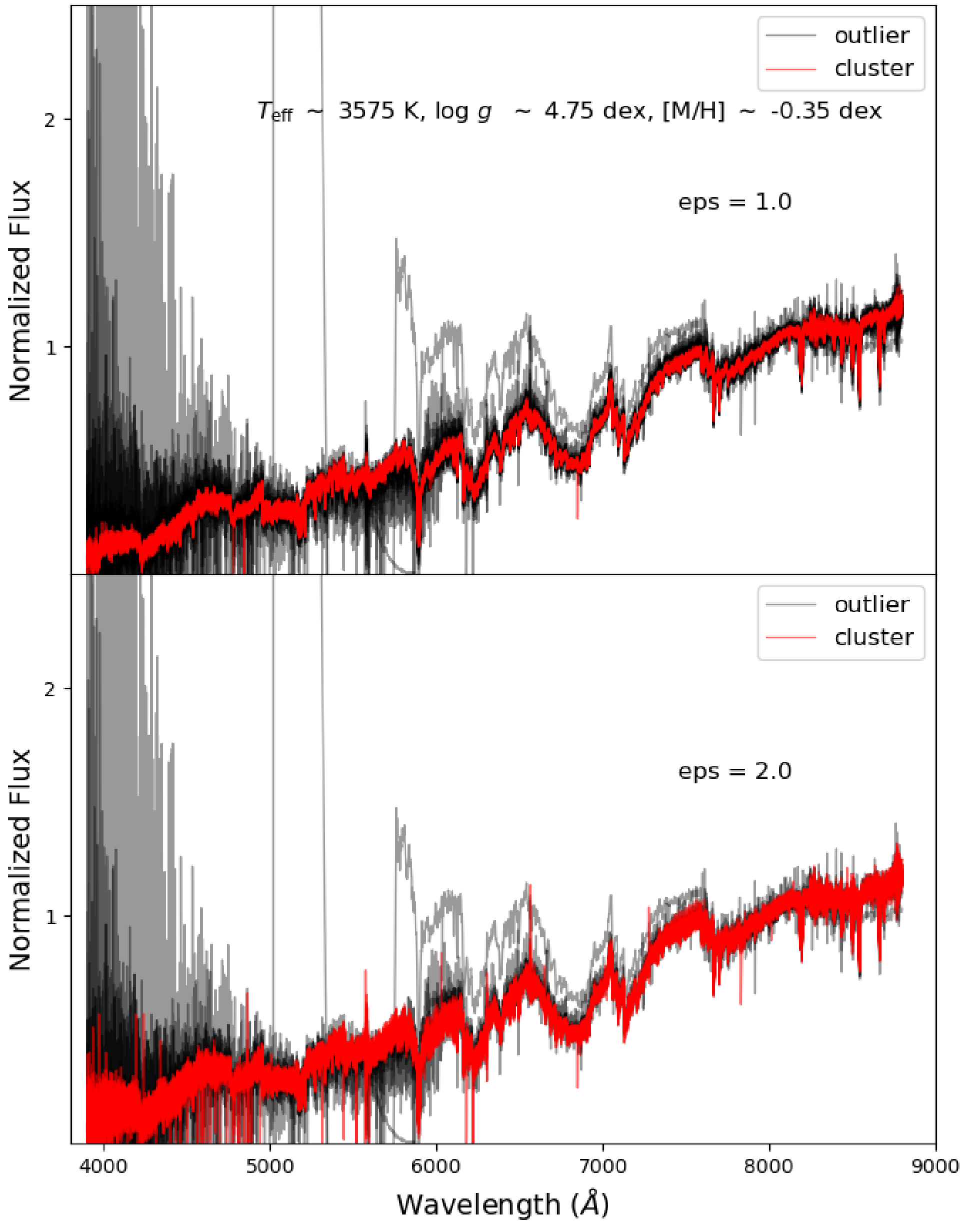

Figure 6. Examples of spectra clusterings for eps values of 1.0 (upper) and 2.0 (lower). The gray lines show all the spectra in the parameter bin of $T_{\rm eff}$: 3550 - 3600 K, $\log g$: 4.7 - 4.8 dex, and [M/H]: -0.3 - -0.4 dex, while the red lines show the spectra that form a cluster.

ing step. We did 10,000 fits for each training. Figure 7 presents the loss function of the last training, indicating that there is no overfit. Figure 8 shows the comparison diagrams of the testing set between the true and predicted values of the parameters. The predicted results were highly consistent with the labeled parameters, with a difference of 14 K for $T_{\rm eff}$, 0.06 dex for $\log g$, and 0.05 dex for [M/H]. This indicates that the parameter labels of the remaining 5105 M-type stars are reliable.

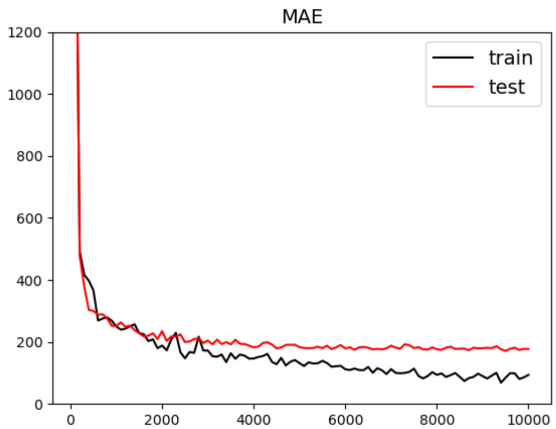

Figure 7. The loss function of the last training with the sampling interval of 100.

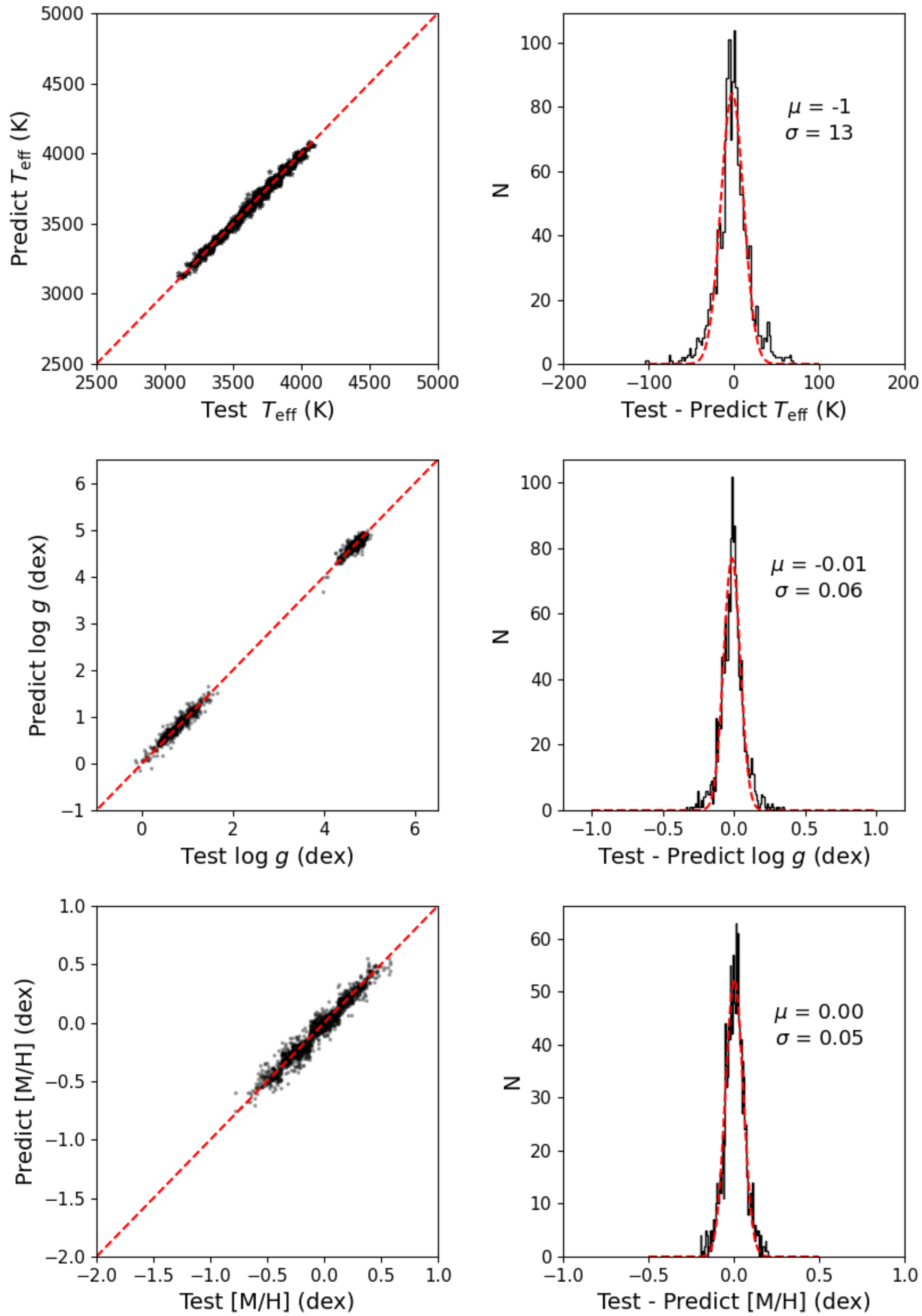

Figure 8. Comparison diagrams of the testing set between the true and predicted values for $T_{\rm eff}$(top), log g(middle) and [M/H] (bottom). The left panel shows one-to-one comparison diagrams and the right panel shows the histograms of differences.

3. VALIDATION AND THE RESULTS

3.1. *Internal Cross-validation*

To verify the reliability of the stellar labels of our samples, we performed an internal cross-validation. We treated each spectrum in the samples as an unknown target and determined its parameters from the remaining samples. We compared these derived stellar parameters to their stellar labels. Figure 9 shows the differences between the derived parameters and their stellar labels, having a scatter of 11 K in $T_{\rm eff}$, 0.05 dex in [M/H], and 0.05 dex in $\log g$. This precision suggests that the stellar labels of our samples are precise.

Our samples span a finit parameter space, and it is essential for targets to match stars in the interior of its reference parameter space. Consequently, the derived parameters of stars at the edge of the parameter space were pulled toward the interior of the parameter distribution, as presented in the right panel of Figure 9. Measurement uncertainties for parameters in sparse samples exceed those in dense samples, and the same holds when these samples serve as a reference set for measuring parameters.

3.2. *The Results*

We ultimately obtained a golden sample that included 5105 M-type spectra with precise stellar labels through multiple cleanings. Their stellar labels consisted of 881 labels from ASPCAP, 1395 from Qiu23, 1651 from Ding22, and 1178 from LASPM. We packaged the stellar parameter catalog and the calibrated spectra in a FITS file, which is available online in the China-VO Paper Data Repository[1].

3.2.1. *The Stellar Parameter Coverage*

Figure 10. shows the Kiel diagram ($T_{\rm eff}$ VS. $\log g$) for the 5105 M-type stars. The PAdova and TRieste Stellar Evolution Code (PARSEC) isochrones at an age of 3 Gyr are also shown in Figure 10. The different colors of the dashed lines represent different [M/H] values from PARSEC version 1.2S (Bressan et al. 2012; Chen et al. 2015). We notice that our samples span a region of the Kiel diagram at $T_{\rm eff}$ $\approx$ 3100-4150 K, $\log g$ $\approx$ -0.3-5.1 dex, and [M/H] $\approx$ -1.0-0.7 dex. We can see separators on Kiel diagram to differentiate metallicities of giant stars, and the separators are consistent with the PARSEC isochrones. Separators to differentiate metallicities are not obvious for dwarf stars. We notice a clear distinction of $\log g$ between giants and dwarfs, and their locations on Kiel diagram are consistent with the PARSEC theoretical tracks.

To ensure the reliability of our samples, we limited our samples to those that have been observed a lot. This causes the parameter space coverage of our samples to be limited in the abundance patterns available within the solar neighborhood. We notice a scarcity of metal-poor stars, which means that the metallicity of a metal-poor star would be overestimated from our samples.

3.2.2. *The Spectra*

A collection of 5105 spectra was acquired, housed within the first Header Data Unit (HDU) of the online FITS file. Each spectrum featured well-calibrated fluxes and wavelengths in the rest frame. Figure 11 displays the spectral comparisons across various temperatures. The SED of the spectrum is notably affected by temperatures, with higher temperatures causing the more distinct features to appear in the blue region of the spectrum. In the red region of the spectrum, the TiO molecular bands near 7050 and 8430 Å exhibit a high sensitivity to $T_{\rm eff}$.

Figure 12 shows the spectral comparisons between dwarf and giant stars for a range of temperatures. Each pair of dwarf and giant stars exhibits identical metallicity and temperature. The molecular bands of giant and dwarf stars show a distinct shape difference in the red region of the spectrum ($\lambda > 6700$ Å). The Na- line pairs at around 7680 Åand 8190 Åexhibit significant changes in their line wings due to pressure broadening, making them good gravity indicators.

Figure 13 shows the differences in the spectra between different metallicities. Each pair of spectra exhibits identical surface gravity and temperature. It is observed that the TiO bands vary with metallicity. This alteration in TiO bands is more susceptible to flux calibration compared to stars of earlier types, which results in the M-type metallicity being more challenging to determine compared to $T_{\rm eff}$ and $\log g$.

[1] https://paperdata.china-vo.org/empirical-lib/golden-sample-M/golden_sample_M_stars.zip

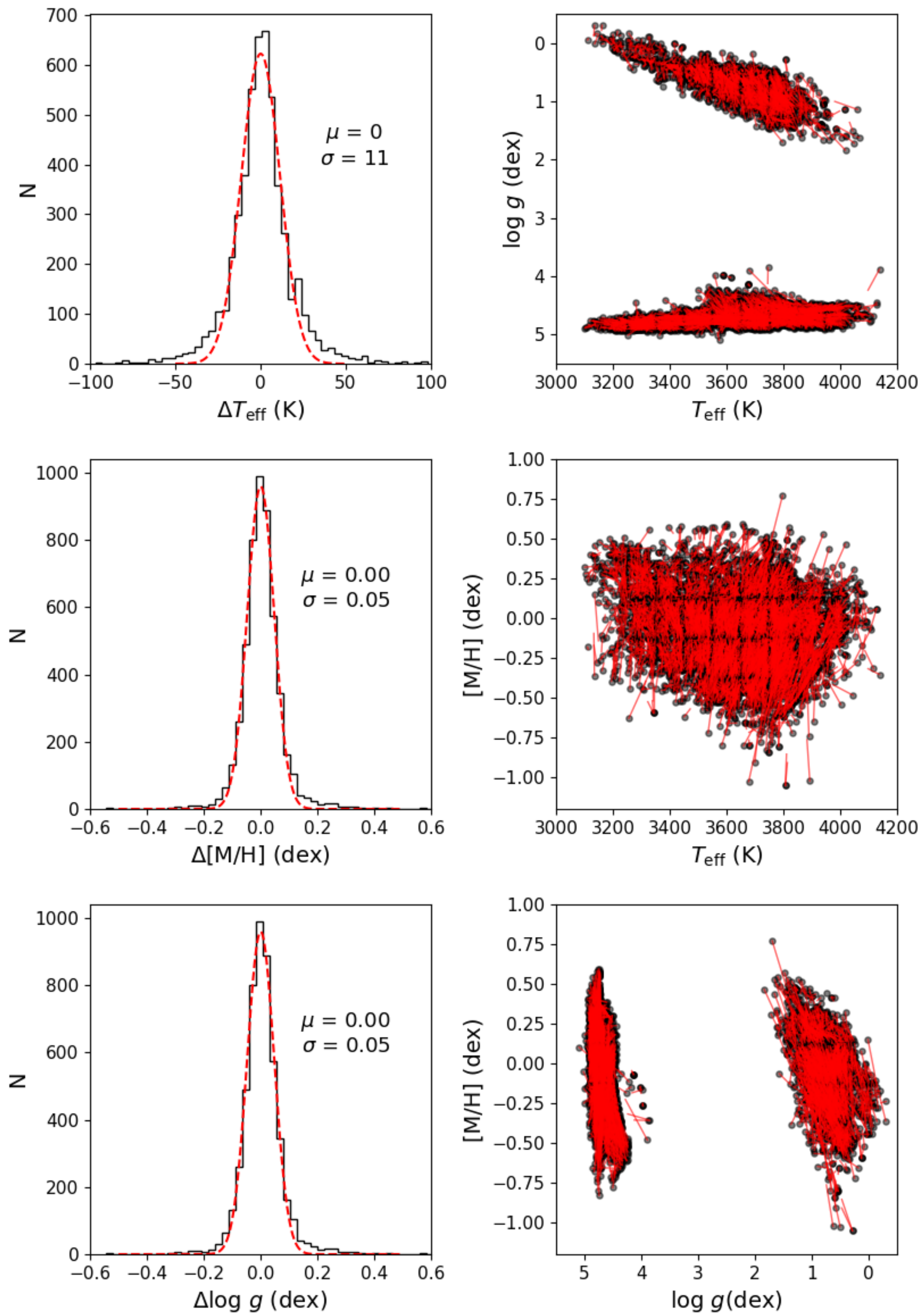

Figure 9. Left panel: histograms of differences between the derived parameters of $T_{\rm eff}$, [M/H], and $\log g$, and their stellar labels. Right panel: black points indicate the stellar labels, while red lines point to the derived parameters.

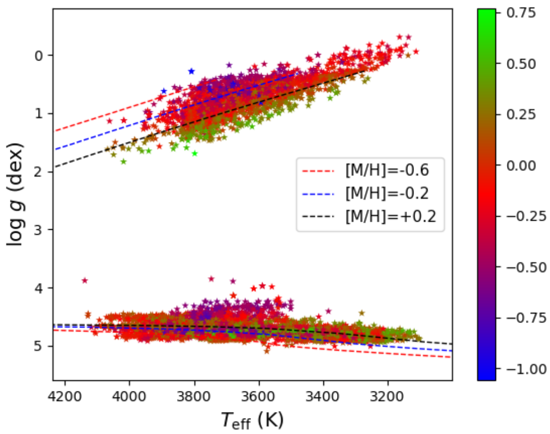

Figure 10. The Kiel diagram of the golden sample. The color is coded by the metallicity [M/H]. The dashed lines represent the isochrones from PARSEC model with different metallicities at the same age of 3 Gyr, i.e., -0.6, -0.2 and 0.2 dex for the red, blue and black lines, respectively.

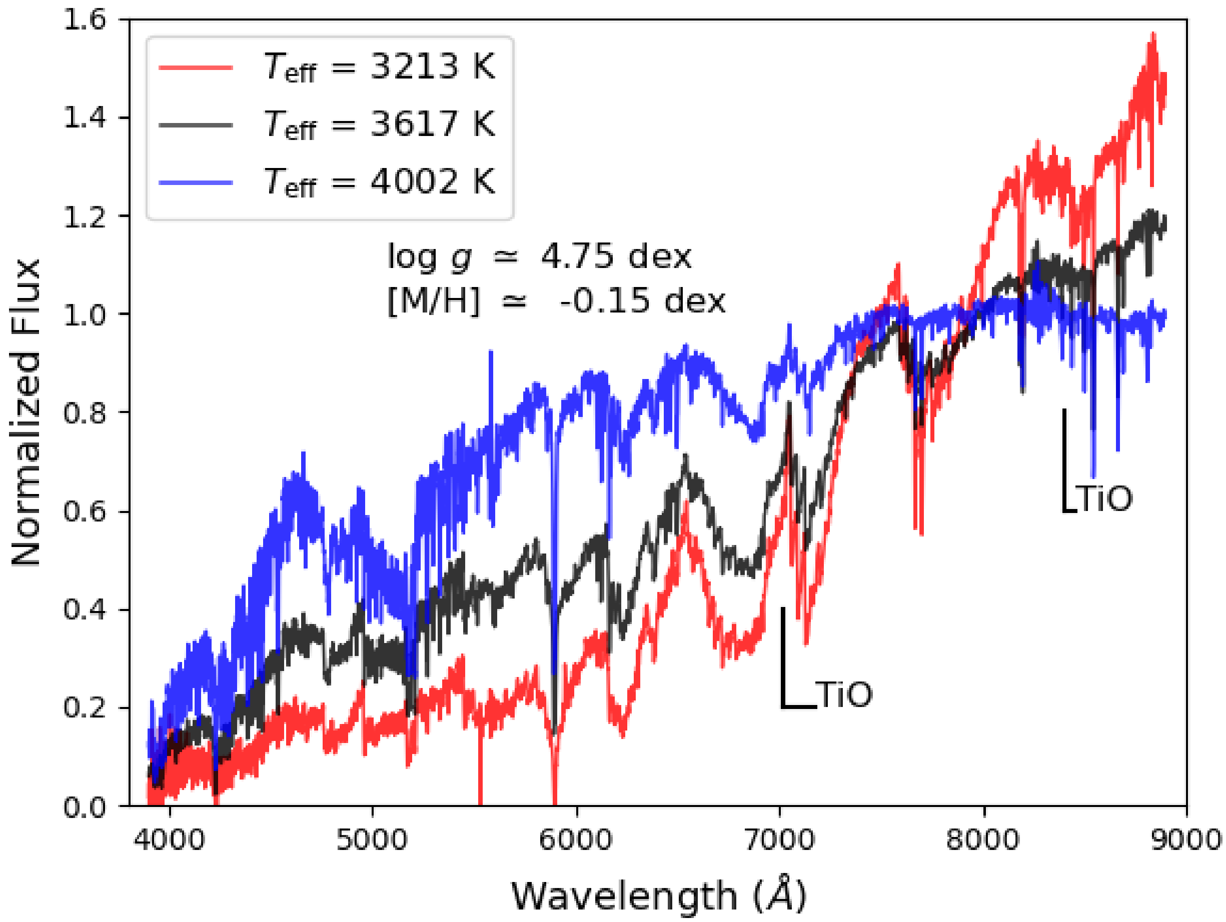

Figure 11. Effective temperature variations for the same surface gravity and Metallicity. The different colors of lines represent different $T_{\rm eff}$ values. The $T_{\rm eff}$-sensitive TiO bands around 7050 and 8430 Å are labeled.

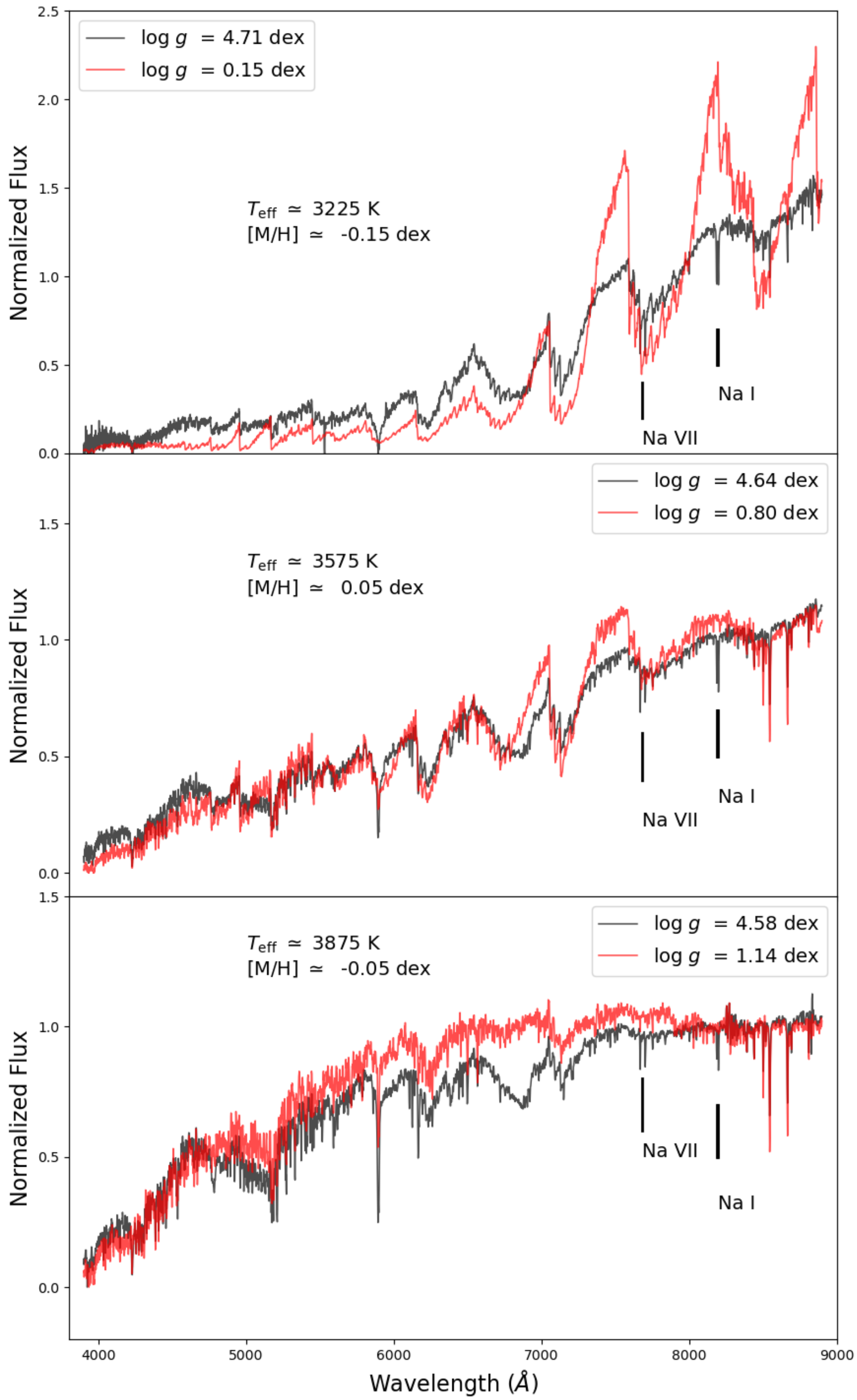

Figure 12. Surface gravity comparison between dwarf and giant spectra of the same metallicity and temperature for the entire spectrum. The black (red) line shows the dwarf (giant) spectrum. The Na- line pairs at around 7680 Åand 8190 Å, which are sensitive to log g, are labeled

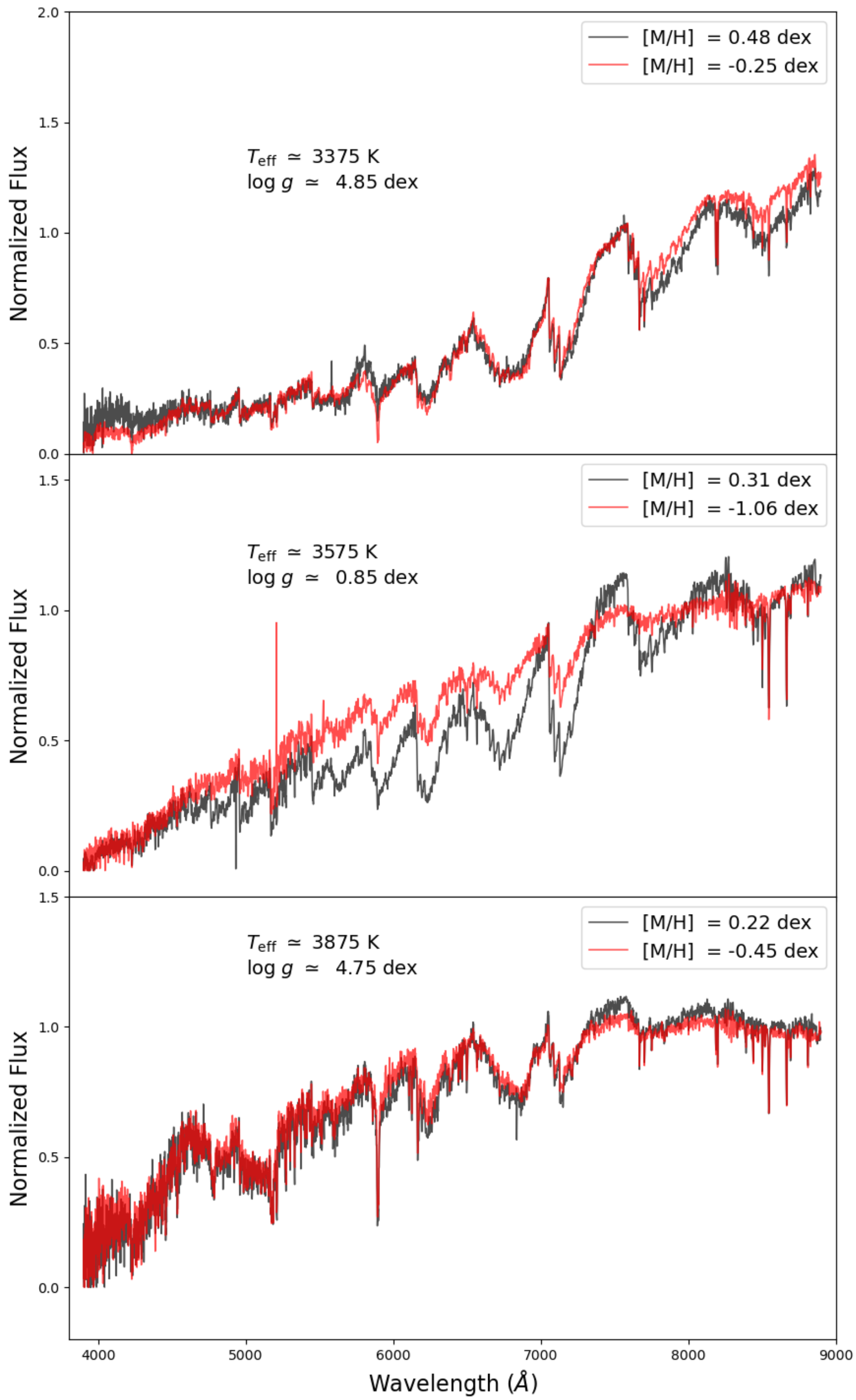

Figure 13. Metallicity comparison between spectra of the same surface gravity and temperature for the entire spectrum.The different colors of lines represent different [M/H] values.

4. APPLICATION TO LASPM

We applied the 5105 samples to LASPM to determine the stellar parameters for 898 350 M-type spectra in LAMOST DR11. The method and the spectral region (6000-8800 Å) we used were the same as described in Du et al. (2021).

4.1. *The Kiel Diagrams*

Figure 14 shows the Kiel diagram for the M-type stars of LAMOST DR11. We also show the PARSEC isochrones as in Figure 10. The stellar parameter distribution of all the M-type stars of DR11 is basically similar to that of the 5105 samples, as expected. Due to the interpolation of the LASPM algorithm, Figure 14 presents a denser parameter space compared to Figure 10, with little gaps being filled by the interpolation.

Figure 15 shows the Kiel diagram for all stars of LAMOST DR11. For AFGK-type stars, the stellar parameters were determined by LASP through comparing the LAMOST spectra to the ELODIE library, which is documented in Luo et al. (2015) and Du et al. (2019). The LASPA (LASP for A-type stars) used different spectral features with those used by LASPFGK (LASP for FGK-type stars), masked the Ca II HK (3900-4060 Å), Hβ (4857-4867 Å), and Hα (6400-6800 Å) to reduce the effect of the features of Am and Ae stars on the parameter measurements (private discussion). Therefore, we can see a small gap around $T_{\rm eff} \sim 8500$ K on the Kiel diagram. Thanks to the StarHorse catalog as a benchmark to correct the systematic differences between the spectroscopic measurements, we ultimately obtained a basically continuous Kiel distribution diagram.

4.2. *Precision of the updated LASPM*

We calculated the parameter precision from the parameter estimates of repeated observations for the same stars. We used an unbiased estimator defined in Du et al. (2021), to measure precision. The estimator is given below:

$$\epsilon = \sqrt{\frac{\rm N}{\rm N-1}} \times (\rm P_i - \overline{P}) \tag{1}$$

where N is the number of times of repeated observations, $\rm P_i$ is the parameter in terms of $T_{\rm eff}$, $\log g$, and [M/H] of the $\rm i_{th}$ observation, and $\rm \overline{P} = \frac{1}{N}\sum_i^N P_i$.

Figure 16 shows the variation of the parameter errors with the spectral S/N. We notice that the parameter errors have a clear increase with decreasing of S/N when S/N < 20, while when S/N > 40, the error is almost constant at $\Delta T_{\rm eff} \approx 60$ K, $\Delta \log g \approx 0.10$ dex and Δ[M/H] ≈ 0.15 dex. Figure 17 shows the Gaussian fits to the ϵ histograms of $T_{\rm eff}$, $\log g$ and [M/H] with spectral S/N $\geq$ 10. The 1σ uncertainties of the ϵ distributions are 67 K for $T_{\rm eff}$, 0.07 dex for $\log g$, and 0.14 dex for [M/H], respectively. The precision (in terms of 1σ uncertainties) of the updated LASPM is improved over its predecessor, which achieved accuracies of 118 K for $T_{\rm eff}$, 0.20 dex for $\log g$, and 0.29 dex for [M/H]. This arises because the reference dataset comprises a dense grid of data that closely mirrors the natural distribution of observations.

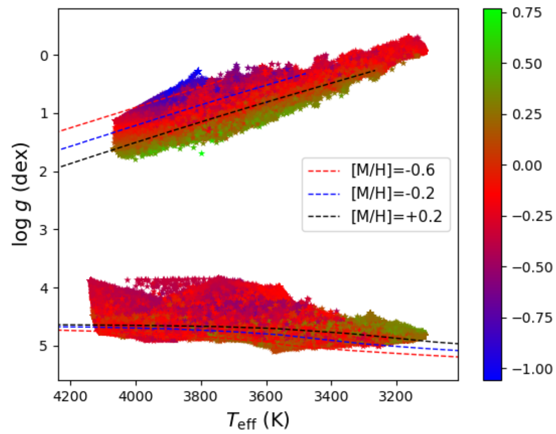

Figure 14. The LASPM-derived Kiel diagram of LAMOST DR11. The color is coded by the metallicity [M/H]. The dashed lines represent the isochrones from PARSEC model with different metallicities at the same age of 3 Gyr, i.e., -0.6, -0.2 and 0.2 dex for the red, blue and black lines, respectively.

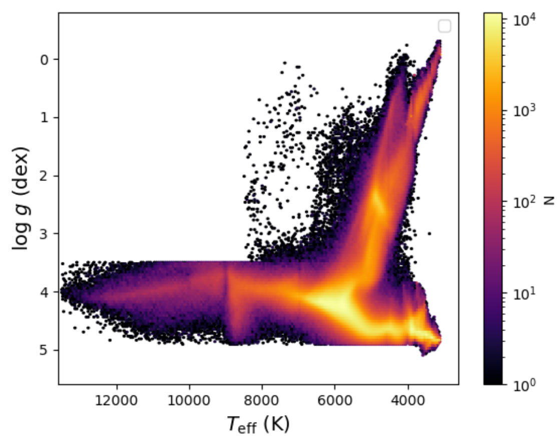

Figure 15. Kiel diagram of all stars of LAMOST DR11, color coded by the density.

4.3. *Comparison with StarHorse*

To investigate whether the LASPM algorithm induces systematic errors in parameter measurements, we compared the LASPM results to the StarHorse catalog. The LASPM parameter catalog was cross-matched with the StarHorse catalog, resulting in a match of 392,509 stars

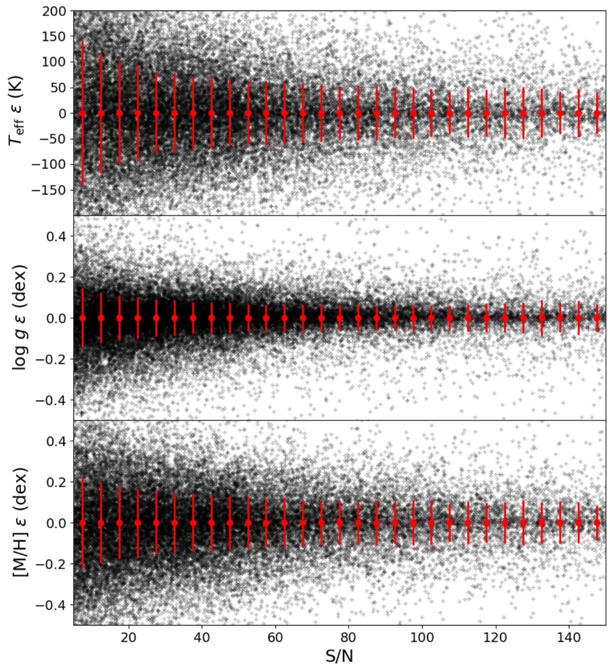

Figure 16. The parameter errors at different S/Ns in terms of $T_{\rm eff}$ (top), $\log g$ (middle), and [M/H] (bottow). The error bars in red represent the parameter errors in different S/N bins, with a step of 5.0.

with the LAMOST S/N >10. Figure 18 shows the comparison of the stellar parameters derived from LASPM with those from StarHorse. The parameters of LASPM are consistent with those of StarHorse, showing a scatter of 103 K in $T_{\rm eff}$, 0.12 dex in $\log g$, and 0.21 dex in [M/H]. The biases of 6 K in $T_{\rm eff}$, 0.02 dex in $\log g$, and 0.03 dex in [M/H] are significantly smaller than the respective scatters, suggesting the absence of systematic offset. The stripe-like overdensities in the bottom left plot correspond to the metallicity resolution of the stellar-model grid that StarHorse used.

4.4. *Comparison with APOGEE*

We cross-matched the LASPM catalog with the APOGEE DR17 following these criteria:

1. STARFLAG = 0;
2. The LAMOST DR11 spectra are identified as M type through the LAMOST 1D pipeline and are parameterized by LASPM;
3. The S/N of the i band for the LAMOST spectra should be larger than 10.

In this way, we selected 12 178 spectra of 6552 M dwarf stars and 3304 spectra of 2068 M giant stars after cross-matching. We adjusted the APOGEE stellar parameters by applying the same offsets between ASPCAP and StarHorse to both dwarfs and giants, respectively. Figure 19 shows the comparison between the parameters of LASPM and the adjusted parameters of ASPCAP for M dwarfs. Likewise, Figure 20 presents the comparison for M giants. The parameters of LASPM matched the ASPCAP parameters fairly well, with a small scatter of 54 K (23 K) in $T_{\rm eff}$, 0.06 dex (0.15 dex) in $\log g$, and 0.12 dex (0.15 dex) in [M/H] for dwarfs (giants). The LASPM $T_{\rm eff}$of the giant is in good agreement with that of ASPCAP, as the stellar labels for all giant stars are either directly or indirectly based on APOGEE data. The $\log g$values for dwarf stars show a strong consistency due to the concentrated distribution of $\log g$among these dwarfs (4.5-5.0 dex).

5. SUMMARY

In this work, we present a golden sample with precise atmospheric parameter labels of M-type stars through stellar label transfer and sample cleaning. The golden sample is available online in FITS format and in the China-VO Paper Data Repository[2]. The golden sample with a balanced parameter distribution, can serve as an empirical library for the determination of stellar parameters and also offers data annotations for data-driven machine learning models. The main work of this paper is summarized as follows.

1. We selected primary samples for M dwarfs and M giants by setting specific distance and magnitude criteria for each. We aligned stellar parameters from various catalogs (SDSS DR17 ASPCAP,Qiu23, Ding22, and LASPM) with the StarHorse catalog to correct systematic discrepancies. We used DBSCAN to remove the unreliable samples in each subgrid of parameters to complete the sample cleaning. Further validation was done using a 5-layer neural network, which confirmed the reliability of the stellar labels for the remaining 5105 stars after outlier removal, showing minimal deviation between predicted and actual values.

2. We conducted an internal cross-validation to assess the reliability of stellar labels of the 5105 stars by treating each spectrum as an unknown and deriving its parameters from the remaining samples. The results showed a scatter of 11 K in $T_{\rm eff}$, 0.05 dex in $\log g$, and 0.05 dex in [M/H], respec-

[2] https://paperdata.china-vo.org/empirical-lib/golden-sample-M/golden_sample_M_stars.zip

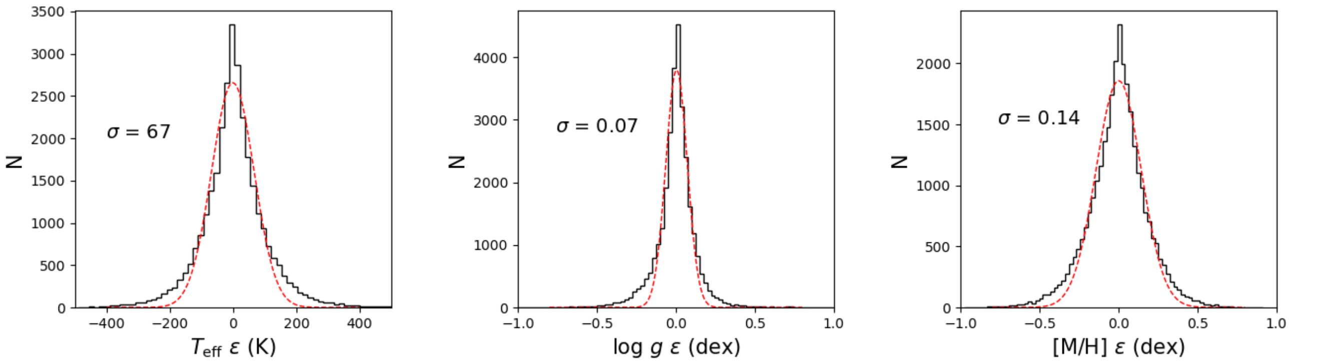

Figure 17. Histograms of ϵ for $T_{\rm eff}$ (left), $\log g$ (middle), and [M/H] (right). The histogram is fitted by Gaussian shown in red dashed curves; The parameter precision for the three parameters are labeled.

tively, indicating a high precision of the stellar labels. A curated collection of 5105 M-type spectra with well-calibrated fluxes and rest-framed wavelengths, each with accurately determined stellar labels from multiple sources, has been compiled into an accessible online fits file.

3. We applied the golden sample to LASPM, producing an almost seamless Kiel distribution diagram for LAMOST DR11 data. The updated LASPM shows improved precision compared to its predecessor, when S/N *geq* 10, with improvements from 118 to 67 K in $T_{\rm eff}$, 0.2 to 0.07 dex in $\log g$, and 0.29 to 0.14 dex in [M/H]. We conducted a comparison of LASPM's parameters with the StarHorse catalog across 392,509 stars. The comparison revealed that the parameters from LASPM matched well with those of StarHorse, exhibiting only slight biases. This suggests that LASPM does not introduce systematic errors in the determination of stellar parameters. The comparison with APOGEE also showed good agreement, confirming the reliability of LASPM in stellar parameter estimation.

By assembling the spectral library from observed spectra instead of theoretical ones, we circumvented the difficulties that current spectral synthesis codes face in accurately depicting the complex spectra of cool stars. This golden sample of M-type stars is very important for many research topics concerning cool stars.

This work is supported by the National Science Foundation of China (grant Nos. xxx and xxx), China Manned Space Project (Nos. CMS-CSST-2021-A10). We thank Mao-Sheng Xiang, Hong-Liang Yan for helpful discussions. Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, the Chinese Academy of Sciences. This research makes use of data from the European Space Agency (ESA) mission Gaia, processed by the Gaia Data Processing and Analysis Consortium.

REFERENCES

Allard, F., Homeier, D., Freytag, B., et al. 2012, EAS Publications Series, 57, 3. doi:10.1051/eas/1257001

Allard, F., Homeier, D., Freytag, B., et al. 2013, Memorie della Societa Astronomica Italiana Supplementi, 24, 128

Anders, F., Khalatyan, A., Chiappini, C., et al. 2019, A&A, 628, A94. doi:10.1051/0004-6361/201935765

Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91. doi:10.1051/0004-6361/202142369

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., et al. 2021, AJ, 161, 147. doi:10.3847/1538-3881/abd806

Bochanski, J. J., Hawley, S. L., Covey, K. R., et al. 2010, AJ, 139, 2679. doi:10.1088/0004-6256/139/6/2679

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127. doi:10.1111/j.1365-2966.2012.21948.x

Chen, X., Wang, S., Deng, L., et al. 2018, ApJS, 237, 28. doi:10.3847/1538-4365/aad32b

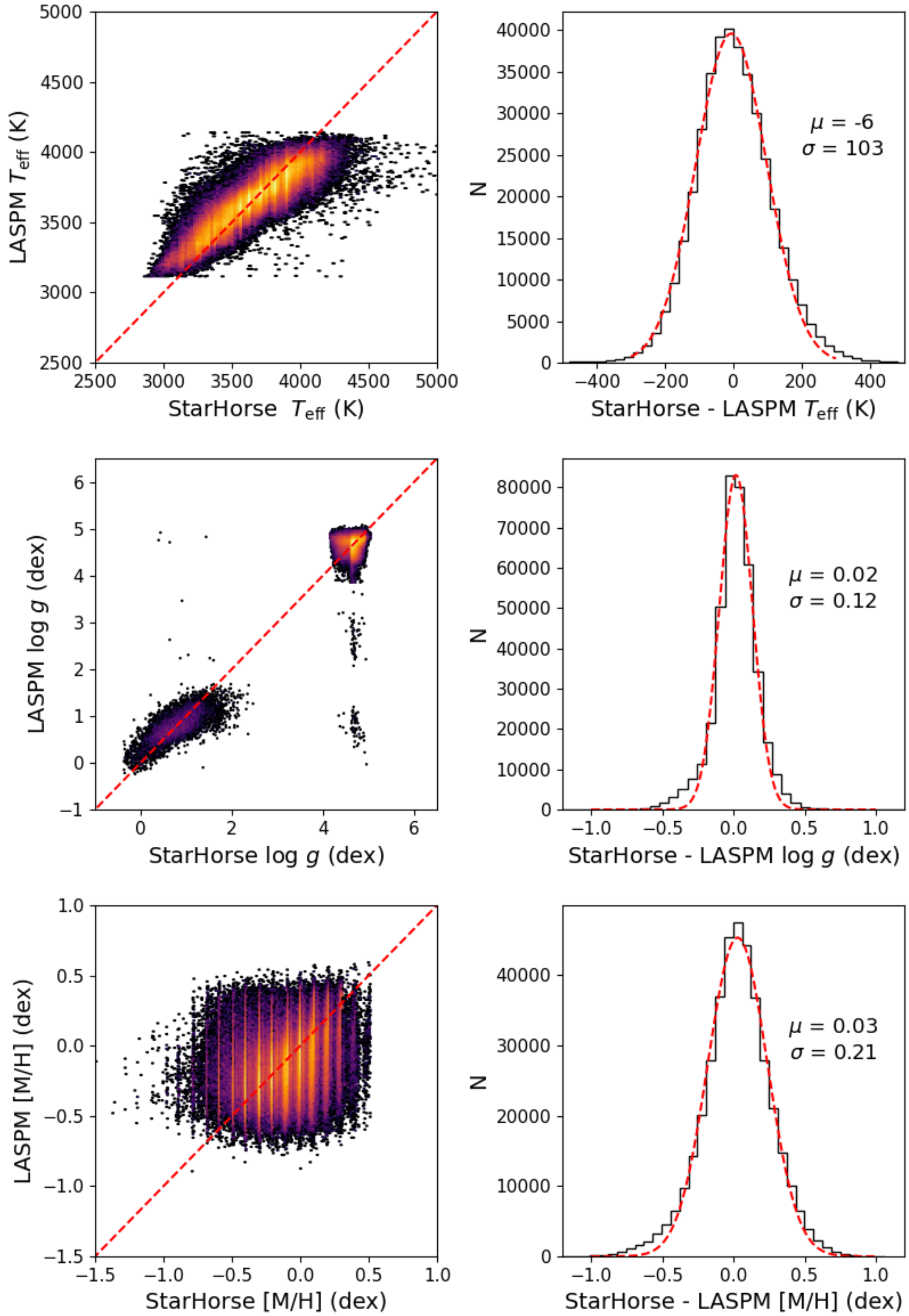

Figure 18. The left panel shows density plots comparing the results from LASPM with those from StarHorse.The right panel shows histograms of differences: no systematic shift was noticed

Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18. doi:10.3847/1538-4365/ab9cae

Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068. doi:10.1093/mnras/stv1281

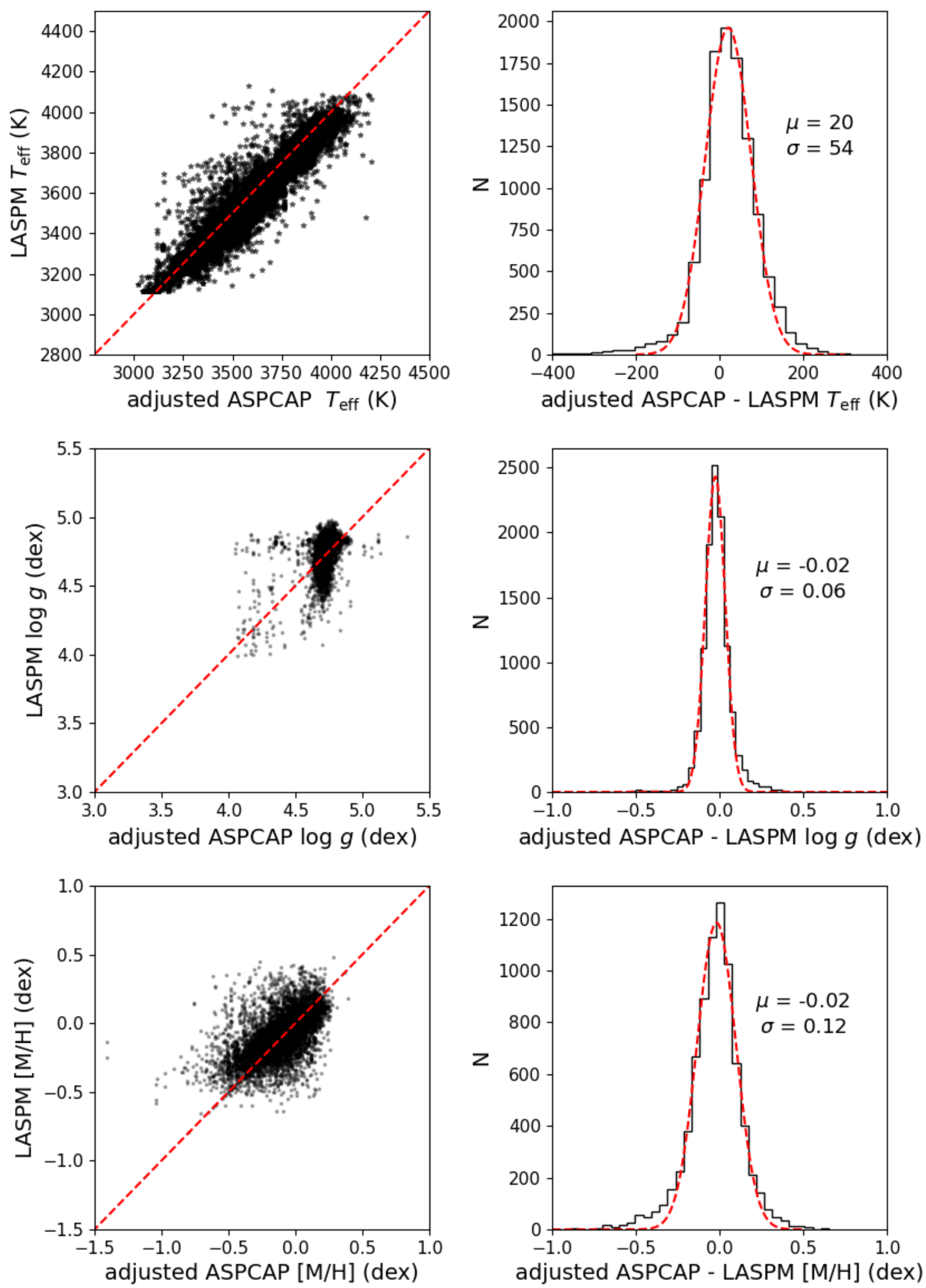

Figure 19. Comparison of the stellar parameters derived from LASPM to the adjusted parameters of ASPCAP for M dwarfs. The left panel shows one-to-one comparison diagrams and the right panel shows the histograms of differences.

Ding, M.-Y., Shi, J.-R., Wu, Y., et al. 2022, ApJS, 260, 45. doi:10.3847/1538-4365/ac6754

Du, B., Luo, A.-L., Zuo, F., et al. 2019, ApJS, 240, 10. doi:10.3847/1538-4365/aaef3c

Du, B., Luo, A.-L., Zhang, S., et al. 2021, Research in Astronomy and Astrophysics, 21, 202. doi:10.1088/1674-4527/21/8/202

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1. doi:10.1051/0004-6361/202039657

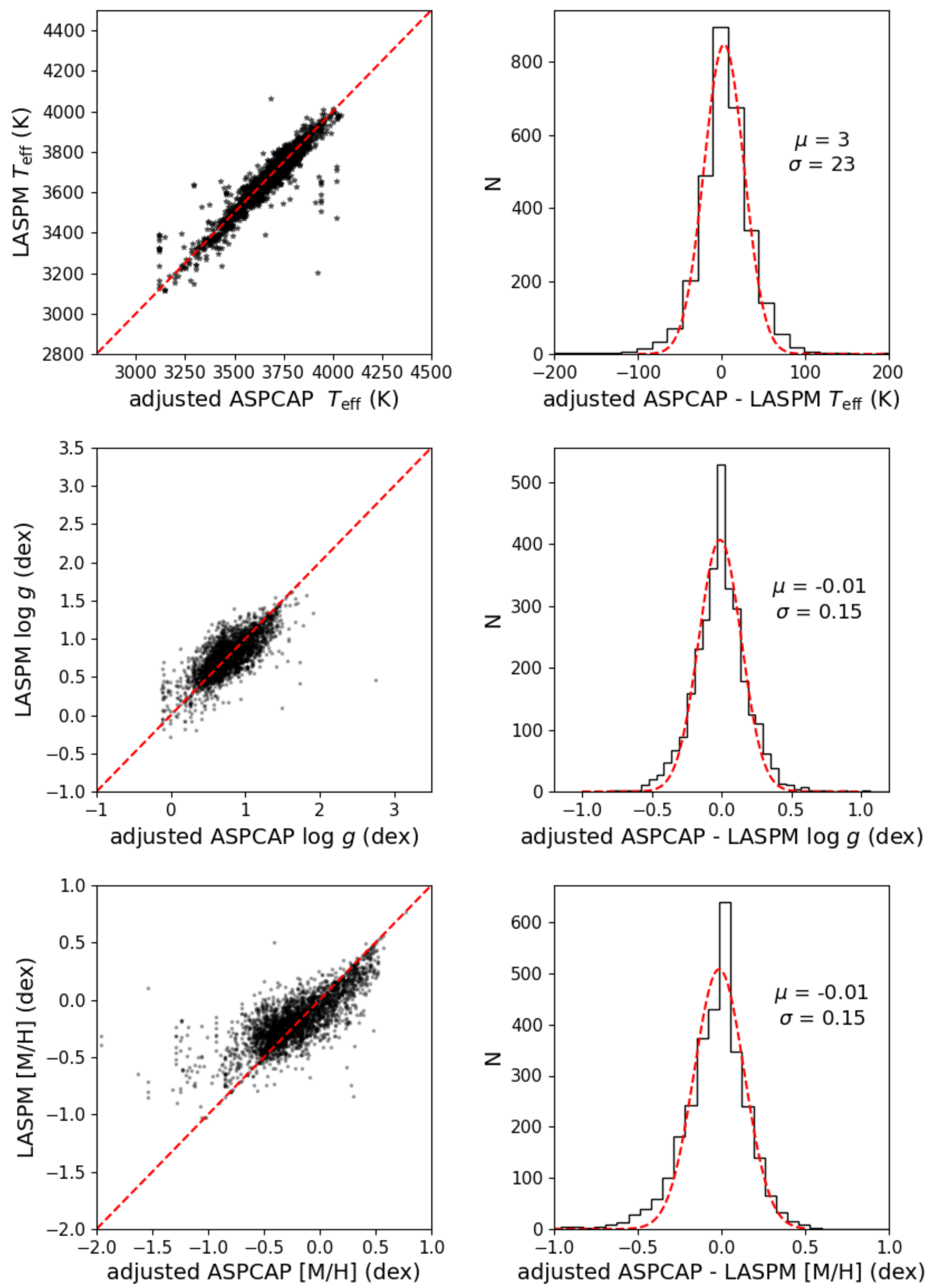

Figure 20. Comparison of the stellar parameters derived from LASPM to the adjusted parameters of ASPCAP for M giants. The left panel shows one-to-one comparison diagrams and the right panel shows the histograms of differences.

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1. doi:10.1051/0004-6361/202243940

Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, MNRAS, 478, 651. doi:10.1093/mnras/sty1008

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951. doi:10.1051/0004-6361:200809724

Howard, E. L., Davenport, J. R. A., & Covey, K. R. 2022, Research Notes of the American Astronomical Society, 6, 96. doi:10.3847/2515-5172/ac6e42

Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, AJ, 160, 120. doi:10.3847/1538-3881/aba592

Kirk, B., Conroy, K., Prša, A., et al. 2016, AJ, 151, 68. doi:10.3847/0004-6256/151/3/68

Li, J., Long, L., Zhong, J., et al. 2023, ApJS, 266, 4. doi:10.3847/1538-4365/acc395

Li, J., Liu, C., Zhang, B., et al. 2021, ApJS, 253, 45. doi:10.3847/1538-4365/abe1c1

Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, Research in Astronomy and Astrophysics, 15, 1095. doi:10.1088/1674-4527/15/8/002

McQuillan, A., Aigrain, S., & Mazeh, T. 2013, MNRAS, 432, 1203. doi:10.1093/mnras/stt536

McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24. doi:10.1088/0067-0049/211/2/24

Passegger, V. M., Wende-von Berg, S., & Reiners, A. 2016, A&A, 587, A19. doi:10.1051/0004-6361/201322261

Plez, B. 2008, Physica Scripta Volume T, 133, 014003. doi:10.1088/0031-8949/2008/T133/014003

Prsa, A., Kochoska, A., Conroy, K. E., et al. 2022, VizieR Online Data Catalog, 225

Qiu, D., Tian, H., Li, J., et al. 2023, Research in Astronomy and Astrophysics, 23, 055008. doi:10.1088/1674-4527/acc153

Qu, C.-X., Luo, A.-L., Wang, R., et al. 2024, ApJS, 270, 32. doi:10.3847/1538-4365/ad103c

Rajpurohit, A. S., Reylé, C., Allard, F., et al. 2013, A&A, 556, A15. doi:10.1051/0004-6361/201321346

Rajpurohit, A. S., Reylé, C., Allard, F., et al. 2014, A&A, 564, A90. doi:10.1051/0004-6361/201322881

Rajpurohit, A. S., Allard, F., Rajpurohit, S., et al. 2018, A&A, 620, A180. doi:10.1051/0004-6361/201833500

Reinhold, T. & Hekker, S. 2020, VizieR Online Data Catalog, 363

Santos, A. R. G., García, R. A., Mathur, S., et al. 2019, ApJS, 244, 21. doi:10.3847/1538-4365/ab3b56

Veyette, M. J., Muirhead, P. S., Mann, A. W., et al. 2017, ApJ, 851, 26. doi:10.3847/1538-4357/aa96aa

Xu, T., Liu, C., Wang, F., et al. 2022, ApJS, 259, 11. doi:10.3847/1538-4365/ac3f2c